Cross-Task Generalization via Natural Language Crowdsourcing Instructions

Abstract

Humans (e.g., crowdworkers) have a remarkable ability in solving different tasks, by simply reading textual instructions that define them and looking at a few examples. Despite the success of the conventional supervised learning on individual datasets, such models often struggle with generalization across tasks (e.g., a question-answering system cannot solve classification tasks). A long-standing challenge in AI is to build a model that learns a new task by understanding the human-readable instructions that define it. To study this, we introduce NATURAL INSTRUCTIONS, a dataset of 61 distinct tasks, their human-authored instructions, and 193k task instances (input-output pairs). The instructions are obtained from crowdsourcing instructions used to create existing NLP datasets and mapped to a unified schema. Using this meta-dataset, we measure cross-task generalization by training models on seen tasks and measuring generalization to the remaining unseen ones. We adopt generative pre-trained language models to encode task-specific instructions along with input and generate task output. Our results indicate that models benefit from instructions when evaluated in terms of generalization to unseen tasks (19% better for models utilizing instructions). These models, however, are far behind an estimated performance upperbound indicating significant room for more progress in this direction.